

Projekt finc

Ein Open Source Discovery System für sächsische Hochschulbibliotheken

von **JENS LAZARUS, BJÖRN MUSCHALL, LEANDER SEIGE** und **EVELYN WEISER**

Kein roter Knopf wurde gedrückt, keine große Rede gehalten: einige Klicks, etwas Tastengeklapper und der neue Katalog der Universitätsbibliothek Leipzig war online. Was am 21. März des Jahres gegen 12 Uhr so unspektakulär und doch nicht ohne Lampenfieber über die Bühne ging, hat inzwischen einige Beachtung gefunden. Es folgten kollegiale Glückwünsche und ein starkes Interesse aus vielen Bibliotheken, von denen die entfernteste sicher die Technical University Sydney ist. Für die Beteiligten aber am wichtigsten war ein weitgehend positives, manchmal geradezu begeistertes Feedback im Chat und im Blog durch die Nutzerinnen und Nutzer der Universitätsbibliothek Leipzig. Die kritischen Hinweise kommen meist aus dem eigenen Haus: Der Katalog ist uns wichtig, wir wollen ihn noch besser machen!

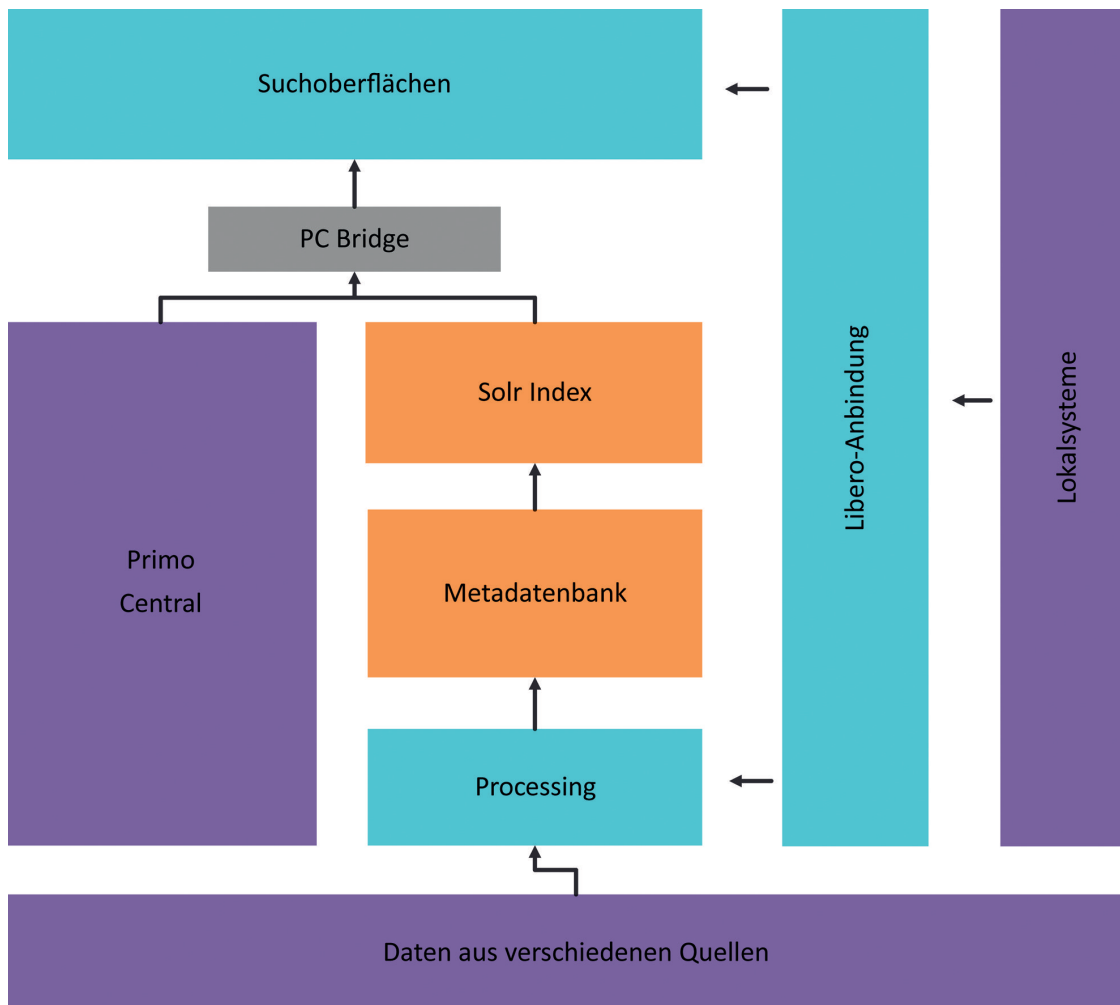
Die neue Oberfläche ist zunächst als sogenannte Beta-Version veröffentlicht. Der in die Jahre gekommene WebOPAC steht noch bis zum Sommer in der zweiten Reihe und dient als Alternative zur Überbrückung von technischen Startschwierigkeiten und als Angebot für Nutzerinnen und Nutzer, die ihre Gewohnheiten nicht ad hoc ändern möchten. Denn jede Neuerung impliziert die Entwöhnung von bisher Vertrauten und eingeübten Abläufen. Der Begriff WebOPAC wird mit der alten Oberfläche verschwinden, die neue heißt in der Universitätsbibliothek Leipzig schlicht „Katalog“. Natürlich unterscheidet sich dieser neue Katalog von einem traditionellen Bibliothekskatalog. Der

WebOPAC bleibt an der Oberfläche: Zeitschriften werden nachgewiesen, Zeitschriftenartikel aus einer Zeitschrift nicht. Der neue Katalog geht hier in die Tiefe. Über einen aggregierten Index, der in die Recherche-Oberfläche integriert ist, werden Zeitschriftenaufsätze und andere Daten im Katalog recherchierbar, die zuvor nur in speziellen Datenbanken zu finden waren. Auch andere externe Datenquellen werden eingebunden. Dass das nicht nur auf gewohnte Publikationsformate beschränkt ist, zeigt sich in der Autorensuche, in welcher zusätzliche Informationen über die Person aus der Wikipedia eingebunden sind. Die Optionen sind umfangreich und öffnen schon quantitativ eine neue Dimension: wurden im WebOPAC etwa drei Millionen Datensätze nachgewiesen, sind es im neuen Katalog weit über 30 Millionen.

Projekt finc

Möglich wurden der neue Katalog, die dahinter liegende technische Infrastruktur und die Anbindung des aggregierten Index durch das aus europäischen Fördermitteln finanzierte und mit Unterstützung des Freistaates Sachsen realisierte Projekt finc (www.finc.info). finc ist dabei ein Kunstwort und kann als find in catalog gelesen werden. Das Projekt finc wird von der Universitätsbibliothek Leipzig koordiniert und hat die Implementierung von modernen Rechercheoberflächen an elf Hochschulen in Sachsen zum Ziel, darunter die Universitäten in Chemnitz, Freiberg und Leipzig, die Fachhochschulen in Dresden, Görlitz / Zittau, Leipzig und

ARCHITEKTUR PROJEKT FINC



Zwickau sowie die Kunst- und Musikhochschulen. Der neue Katalog der Universitätsbibliothek Leipzig war hier nur der erste Meilenstein.

Rückblende

Nach intensiven Diskussionen um eine moderne Katalogoberfläche wurde hierfür im Jahr 2010 eine Interessengemeinschaft sächsischer Hochschulbibliotheken unter der Regie der Universitätsbibliothek Leipzig gegründet. Diese hatte das Ziel, einen EFRE-Projektantrag für eine indexbasierte Rechercheoberfläche auf den Weg zu bringen und eine gemeinsame Lösung zu realisieren. Bereits im Frühjahr 2010 hatte sich die SLUB Dresden, wie später auch die Hochschule Mittweida, für den Einsatz von Primo der Firma ExLibris entschieden.

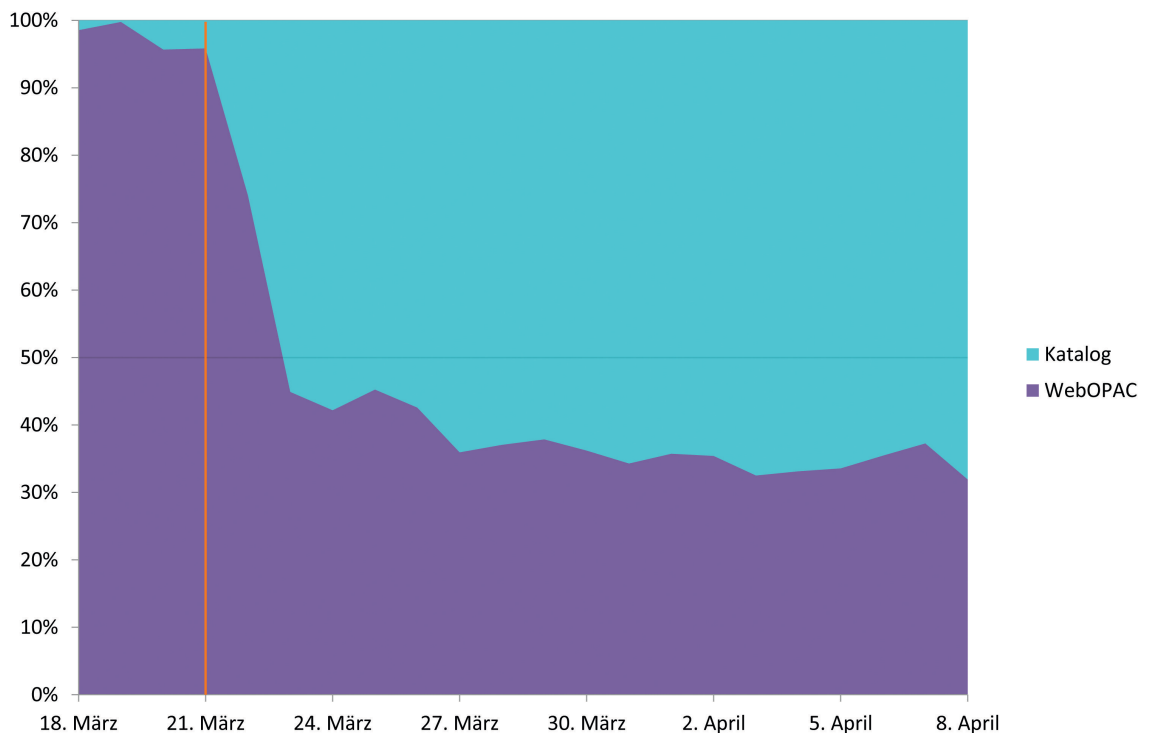
Die Suche nach Lösungen lief in zwei Richtungen. Zum einen erfolgte die Prüfung kommerzieller Produkte. Zum anderen wurde aber auch intensiv die Eignung von Open Source Software evaluiert. Um hier zu belastbaren Einschätzungen zu kommen und die notwendige Kompetenz aufzubauen, testete ein dreiköpfiges Projektteam aus der Universitätsbibliothek Leipzig (Björn Muschall, Leander Seige) und

der Bibliothek der Hochschule für Musik und Theater in Leipzig (Anke Hofmann) die quelloffene Software VuFind (www.vufind.org). Innerhalb weniger Wochen konnte das Team mit einer überzeugenden Machbarkeitsstudie die Lösung für die als besonders schwierig angenommenen Aufgabenstellungen nachweisen.

Die Weichenstellung der Interessengemeinschaft für eine Open Source Software war letztlich eine strategische Entscheidung. Die Katalogoberfläche ist das wichtigste Werkzeug von Bibliotheken, die eigenen Ressourcen anzubieten und zu vermitteln. Für die beteiligten Einrichtungen wogen die hohen Freiheitsgrade, die Möglichkeit der selbstbestimmten Entwicklung, die Flexibilität in der künftigen Einbindung externer Dienstleistungen und Informationsquellen die Nachteile einer community-abhängigen, freien Software auf.

Am 17. März 2011 wurde der EFRE-Antrag zur Implementierung einer suchmaschinenbasierten Rechercheoberfläche an Hochschulbibliotheken in Sachsen eingereicht. Durch einen vorzeitigen Maßnahmebeginn konnte ab Jahresmitte die Suche nach

NUTZERBEWEGUNG VOM WEBOPAC ZUM NEUEN KATALOG



geeigneten Projektmitarbeitern vorangetrieben werden. Die Stellenbesetzungen erfolgten ab September 2011. Die Arbeit konnte beginnen. Am 7. November 2011 wurde der EFRE-Antrag mit einem Fördervolumen von 1,3 Millionen Euro und einer Laufzeit über drei Jahre bewilligt.

Projektstart

In den ersten Wochen des Projektes wurde die Basis für den konsortialen Betrieb geschaffen. Neben der Beschaffung und dem Aufbau der Hardware wurde ein Projektmanagementsystem eingerichtet, welches die einzelnen Entwicklungsschritte und deren Bearbeitungsstand abbildet und eine effiziente Zusammenarbeit aller beteiligten Bibliotheken ermöglicht. Ein Auftaktmeeting fand im Oktober 2011 statt. Es folgten Workshops in Leipzig, Chemnitz und Freiberg, um die Entwicklungen zusammen mit den Kollegen vor Ort voranzubringen.

Stichwort: Aggregierter Index

Bestandteil des Projektes ist neben dem sogenannten Discovery Service – der Katalogoberfläche mit der dazugehörigen Infrastruktur – auch die Anbindung eines aggregierten Index. In einem solchen Index werden Datensätze aus einer Vielzahl von Datenbanken verschiedener Anbieter und Verlage zusammengeführt (aggregiert). Dieser Mega-Index kann mehrere hundert Millionen Datensätze umfassen und als eigenständiges Produkt lizenziert werden. Nach einer europaweiten Ausschreibung und einem intensiven Evaluierungsprozess auf der Basis eines

umfangreichen Kriterienkataloges und einer Teststellung in realer Umgebung fiel die Entscheidung auf den aggregierten Index Primo Central der Firma ExLibris, den auch die SLUB im Rahmen ihrer Auswahlentscheidung seit Ende 2010 einsetzt.

Workstattberichte

Das Hosting der Suchmaschinen, Datenbanken und Webserver für alle beteiligten Bibliotheken wird an der Universitätsbibliothek Leipzig realisiert. Hier laufen auch die Verbindungen zu den Systemen der Projektpartner zusammen.

Nico Törl, Unix-Systemadministrator: Der Projektstart verlief anfangs etwas holprig. Da die benötigte Hardware nicht zur Verfügung stand, begannen wir das Projekt mit einer Interimslösung. Zur Anbindung der individuellen Libero-Instanzen benötigen wir gesicherte Verbindungen zu den Bibliothekssystemen der Bibliotheken, wobei die verschiedenen Häuser unterschiedliche lokale Gegebenheiten und Bedürfnisse haben, denen wir bestmöglich entgegenkommen.

Unsere Nutzer und auch die Bibliotheksmitarbeiter geben uns großartiges Feedback zu möglichen Verbesserungen. Wir versuchen, diese Vorschläge so schnell wie möglich umzusetzen, was insbesondere bei strukturellen Änderungen aufgrund der enormen Datenmengen nicht immer ganz einfach ist. Die technische Grundlage des neuen Kataloges ist die Open Source Software VuFind. Diese hat ihren

Ursprung an der Villanova University (Pennsylvania, USA) und wird von einer internationalen Gemeinschaft von Softwareentwicklern gepflegt. Das Projektteam hat die spannende Aufgabe, die Software an die hiesigen Verhältnisse anzupassen.

Eine besondere Herausforderung stellt die Anbindung der Bibliothekssysteme dar. Bestimmte Informationen müssen in Echtzeit in den Katalog übertragen werden, zum Beispiel ob ein Buch gerade verfügbar oder ausgeliehen ist. Hierfür wurde eine spezielle Software entwickelt, welche die Zugänge zu den Bibliothekssystemen bündelt und über eine einheitliche Schnittstelle den Discovery Systemen zugänglich macht.

POLICHRONIS TSOLAKIDIS, Java-Programmierer: Die Anbindung an Libero erfolgt über eine JDBC-Schnittstelle sowie Webscraping der WebOPAC-Oberfläche. Nach außen kommuniziert die Software als Webservice, welche die ermittelten Daten im JSON-Format bereitstellt. Zugleich unterstützt die Software das Metadatenprocessing.

Auf Grund des Alters von Libero/WebOPAC gestaltet sich die Anbindung nicht einfach. Der HTML-Code ist nicht konform mit den W3C-Vorgaben. Die Syntax ist ziemlich problematisch. Bei Javascript verlässt sich das Programm ebenfalls auf den Langmut des Browsers. Die Interaktion zwischen WebOPAC und HTTP-Client verläuft dagegen problemlos.

Als Datenbank kommt bei Libero InterSystems Caché zum Einsatz, bei welcher die JDBC-Schnittstelle eher ein Alibi darstellt. In Zukunft wird die Nutzung von Open Source die bessere Wahl sein, um auf die schnellen Veränderungen, gerade im medialen Bereich, in angemessener Zeit zu reagieren.

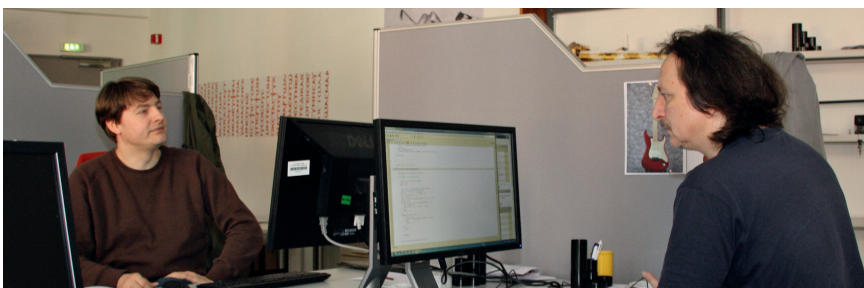
Die Titel- oder Metadaten zu den Beständen der Bibliotheken werden über den Südwestdeutschen

Bibliotheksverbund (SWB), im Format MARC21 bereitgestellt und in den Index der Suchmaschine übernommen. Diese Daten werden um Informationen aus dem jeweiligen Bibliothekssystem ergänzt. Neben den Titeldaten der Bibliotheken werden auch die Daten aus zahlreichen externen Quellen eingespielt, die in unterschiedlichen Formaten vorliegen.

EVELYN WEISER, System-Bibliothekarin: Um in den Daten die spezifischen Informationen der einzelnen Häuser abzubilden, haben wir das bibliothekarische Austauschformat MARC sehr stark erweitert. Es gibt da Kategorienbereiche, die lokalen Informationen vorbehalten sind. So können wir bibliotheksspezifische Daten wie Signaturen oder Standorte gemeinsam mit den normalen Titeldaten speichern und für die Indexierung und Anzeige aufbereiten. MARC ist ein sehr komplexes und nicht mehr ganz zeitgemäßes Format, das sicher viele Wünsche offen lässt – in unserem Projekt allerdings hat es sich bisher als durchaus brauchbar und erstaunlich flexibel herausgestellt. Um das Ausgangsformat von dem zu unterscheiden, was am Ende indexiert wird, haben wir einen neuen Begriff geprägt – wir sprechen vom sogenannten fincMARC.

Die neuen Kataloge bieten die Möglichkeit, über die jeweiligen Bibliotheksbestände hinaus weitere Quellen unter einer Oberfläche verfügbar zu machen. Zu diesem Zweck werden wir zahlreiche zusätzliche Datenmengen verarbeiten und in die Suchmaschine indexieren, darunter Inhalte des Project Gutenberg, verschiedene Hochschulschriftensammlungen oder auch sehr spezielle musikwissenschaftliche Datenbanken. Jede Bibliothek kann entscheiden, welche dieser Sammlungen sie neben den eigenen Beständen zur Verfügung stellen möchte.

Die Aufbereitung und zentrale Speicherung der Importdaten erfolgt in einer klassischen, relationalen Datenbank. Diese zu füllen ist der Gegenstand



der zweiten zusätzlichen Software, die im Projekt neu entwickelt wird.

MARTIN CZYGAN, Programmierer: Zwei Herausforderungen stellen sich bei der Vorverarbeitung der bibliothekarischen Metadaten: Wir müssen eine Flexibilität herstellen und erhalten, die es erlaubt, Metadaten unterschiedlicher Herkunft in einem einheitlichen Prozess zu behandeln. Bibliographische Formate sind nicht trivial und jedes einzelne besitzt seine Eigenheiten – um die Komplexität nicht weiter durch eine komplizierte Software zu erhöhen, streben wir eine modulare und kleine Lösung an. Das Wichtigste (auch für den Nutzer) sind die Daten; die Software sollte eine dünne Schicht bilden, um diese Daten zu reorganisieren und anzureichern – nicht mehr.

Gleichzeitig wollen wir in der Lage sein, die uns zur Verfügung stehenden Datenmengen schnell zu verarbeiten. In den letzten Wochen sind wir dem Ziel einer massiv-parallelen Architektur näher gekommen und haben die Verarbeitungszeiten drastisch verkürzen können. Und wir sind zuversichtlich, durch Einsatz von weiteren Optimierungstechniken wie Caching diese noch weiter senken zu können. Das Ziel, neben all den technisch interessanten Herausforderungen, bleibt natürlich, das Informationsangebot im Sinne des Bibliotheksnutzers zu verbessern.

Die Daten aus allen angeschlossenen Systemen laufen in der Katalogoberfläche, dem sogenannten Frontend zusammen; Daten aus der Suchmaschine, dem aggregierten Index, den Bibliothekssystemen und externen Diensten werden miteinander verknüpft und in Funktion gebracht.

FRANK MORGNER, Assistent des Projektmanagements und Programmierer: VuFind bringt alle grundlegende Funktionalitäten eines Discovery Systems mit. Die sächsischen Hochschulbibliotheken nutzen LIBERO als Bibliothekssystem, und so geht es außerdem darum, auch dessen Funktionalitäten möglichst in Gänze zu übernehmen. Hierzu mussten wir allerdings VuFind um eine Schnittstellenanbindung an das Bibliothekssystem erweitern. Dabei ist es aus unserer Sicht wichtig, immer wieder einen guten Kompromiss zu finden, um VuFind kompatibel und konfigurierbar zu halten. Große Abweichungen zum ursprünglichen VuFind Quellcode versuchen wir zu vermeiden. Eine zusätzliche Herausforderung ist sicherlich die Betreuung von elf Bibliotheken. Diese müssen zwar die Möglichkeit haben, ihre Katalog-Oberflächen nach eigenen Vorstellungen zu gestalten, gleichzeitig soll jedoch der logische VuFind-Quellcode weiterhin erhalten bleiben.

Feedback

Seit dem 21. März 2012 ist der neue Katalog an der Universitätsbibliothek Leipzig online. Auf verschiedenen Wegen wird das Feedback der Nutzerinnen und Nutzer „eingesammelt“ und ausgewertet. Neben

Interviews und den Kontakten an den Beratungs- und Ausleihtheken gab es zahlreiche Kommentare im Blog zum Katalog. Auch im Chat oder in den E-Mail-Anfragen ist der neue Katalog immer wieder Thema. Es handelt sich nahezu ausnahmslos um freundliche, konstruktive Kritik oder gar Lob. Das Ausbleiben der großen Flut an Rückmeldungen ist vorsichtig positiv zu bewerten. Bei negativen Erfahrungen ist die Reaktionsbereitschaft in der Regel höher als bei unproblematischen oder gar positiven Eindrücken. Besonders aufmerksam wurden Reaktionen auf die Integration des Primo Central Index ausgewertet. Wie würden Nutzerinnen und Nutzer mit der Recherche in einer um ein zig-faches gestiegenen Titelmenge zurechtkommen? Gerade diese Möglichkeit erfährt aber sehr große Zustimmung.

Die Universitätsbibliothek Leipzig kommt als erste Hochschulbibliothek Sachsens in den Genuss eines fine-Kataloges. Sie hat dabei aber zwangsläufig die Engpässe und Kinderkrankheiten in der Startphase eines solchen Projektes mit zu tragen. Die Rückmeldungen der Nutzerinnen und Nutzer dienen der Universitätsbibliothek Leipzig zur Evaluierung ihres neuen Kataloges und werden selbstverständlich auch zur Optimierung von Funktionalität und Oberfläche herangezogen.

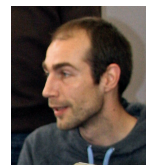
Ausblick

In der Beta-Version geht es vorrangig um stabile Grundfunktionalitäten des Kataloges. Ist das erreicht, kann die Gestaltung und Entwicklung weiter vorangetrieben werden. So wird es im Rahmen einer Masterarbeit eine Usability-Studie zum Katalog geben. Während der Projektlaufzeit sollen außerdem die bereits vorhandenen Normdaten angereichert werden, um auch fremdsprachliche Ansetzungsformen von Personennamen zu berücksichtigen. Auch ist die Entwicklung von Anwendungen für mobile Endgeräte und die Anbindung weiterer Wissensdatenbanken vorgesehen.

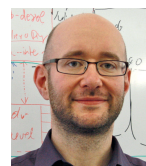
In den kommenden Monaten werden schrittweise alle beteiligten sächsischen Universitäts- und Hochschulbibliotheken mit der neuen Recherche-Oberfläche ausgestattet. Die Bibliotheken der Hochschule für Musik und Theater Leipzig und der Universitäten in Chemnitz und Freiberg werden in der Jahresmitte ihre neuen Kataloge einführen. Die gemeinsame Entwicklung in Sachsen ist eine spannende Herausforderung und wird über die Projektlaufzeit hinausgehen. Der Start an der Universitätsbibliothek Leipzig jedenfalls scheint gelungen.



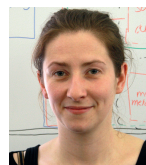
JENS
LAZARUS



BJÖRN
MUSCHALL



LEANDER
SEIGE



EVELYN
WEISER